ARI Research Note 89-22

# Effects of Stress on Judgment and Decision Making in Dynamic Tasks

Kenneth R. Hammond
University of Chicago

for

Contracting Officer's Representative
Judith Orasanu

Office of Basic Research
Michael Kaplan, Director

May 1989



United States Army
Research Institute for the Behavioral and Social Sciences

# U.S. ARMY RESEARCH INSTITUTE
# FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel**



---

Research accomplished under contract
for the Department of the Army

University of Colorado

Technical review by

Judith Orasanu

**NOTICES**

**DISTRIBUTION:** This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Informational Service (NTIS).

**FINAL DISPOSITION:** This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

**NOTE:** The views, opinions, and findings in this report are those of the author(s) and should not to be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

| | |
|---|---|
| 1a. REPORT SECURITY CLASSIFICATION<br>Unclassified | 1b. RESTRICTIVE MARKINGS<br>-- |
| 2a. SECURITY CLASSIFICATION AUTHORITY<br>--<br>2b. DECLASSIFICATION/DOWNGRADING SCHEDULE<br>-- | 3. DISTRIBUTION/AVAILABILITY OF REPORT<br>Approved for public release; distribution is unlimited. |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S)<br>-- | 5. MONITORING ORGANIZATION REPORT NUMBER(S)<br>ARI Research Note 89-22 |
| 6a. NAME OF PERFORMING ORGANIZATION<br>University of Colorado<br>6b. OFFICE SYMBOL (If applicable)<br>-- | 7a. NAME OF MONITORING ORGANIZATION<br>U.S. Army Research Institute<br>Office of Basic Research |
| 6c. ADDRESS (City, State, and ZIP Code)<br>Center for Research on Judgment and Policy<br>Campus Box 344 - University of Colorado<br>Boulder, CO 80309-0344 | 7b. ADDRESS (City, State, and ZIP Code)<br>5001 Eisenhower Avenue<br>Alexandria, VA 22333-5600 |
| 8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences<br>8b. OFFICE SYMBOL (If applicable)<br>PERI-BR | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br>MDA903-86-C-0142 |
| 8c. ADDRESS (City, State, and ZIP Code)<br>5001 Eisenhower Avenue<br>Alexandria, VA 22333-5600 | 10. SOURCE OF FUNDING NUMBERS<br>PROGRAM ELEMENT NO. 61102B<br>PROJECT NO. 74F<br>TASK NO.<br>WORK UNIT ACCESSION NO. |

11. TITLE (Include Security Classification)

Effects of Stress on Judgment and Decision Making in Dynamic Tasks

12. PERSONAL AUTHOR(S)

Hammond, Kenneth R. (University of Colorado)

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT (Year, Month, Day) | 15. PAGE COUNT |
|---|---|---|---|
| Interim | FROM 87/09 TO 88/08 | 1989, May | 37 |

16. SUPPLEMENTARY NOTATION

| 17. COSATI CODES | | | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Dynamic tasks, Uncertainty, Experts, Judgment |

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Studies of expert microburst forecasters were conducted. Two studies yielded results confirming the validity of a linear model of expert judgment and the meaningfulness of profiles as representations of weather phenomena. A simulation demonstrated that a simple scientifically and empirically ignorant forecasting model could perform as well as a sophisticated scientifically informed algorithm. A study conducted under dynamic and highly representative forecasting conditions yielded the following major findings: (a) agreement on precursor values was low to moderate, setting the possible upper limit on forecasting accuracy; (b) agreement on microburst probabilities was lower under the highly representative situation than in our "best case scenario" study; and (c) new information received over time had very little impact on the experts' judgments.

| | |
|---|---|
| 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT<br>☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS | 21. ABSTRACT SECURITY CLASSIFICATION<br>Unclassified |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL<br>Judith Orasanu | 22b. TELEPHONE (Include Area Code) (202) 274-8722<br>22c. OFFICE SYMBOL PERI-BR |

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

# EFFECTS OF STRESS ON JUDGMENT AND DECISION MAKING IN DYNAMIC TASKS

## Goals

The principal goals of this research project are to (a) discover the nature of judgment and decision making in dynamic tasks and (b) study the effects of stress on judgment and decision making in such tasks. Neither project has been undertaken previously by researchers in this field.

The research carried out is best summarized in the report of May 1988 in which the results of six studies of expert decision making in dynamic tasks are described. The six studies are cumulative, that is, each study is a logical consequence of its predecessor; the final study (VI) is of greatest importance because it involves experts making judgments in a dynamic task situation that is highly representative of their working conditions. The results thus obtained carry an authenticity not ordinarily available. A report describing Study VI (Appendix A) and a summary of the studies preceding it together with research recommendations (Appendix B) are appended to this report.

Since May 1988 we have made further analyses of the probability judgments of the experts regarding microburst events. This work was described in the eighth Quarterly Report (see Appendix C). The most significant findings were that (a) new information received over time had very little impact on the experts' judgments and (b) the experts were very poorly coordinated with one another.

Thus we found a group of five experts working together on an important problem for several years, yet never comparing their performances. We also found that well-known psychological research procedures produced information heretofore unknown--and unsought--that was fundamental to that research effort (detecting and forecasting microbursts). Our research on hail forecasting produced similar results. In short, the most significant results of our work are that (a) important information regarding expert judgment in dynamic tasks can be produced rapidly with standard research techniques, (b) that such information will not be produced unless and until such techniques are applied, and (c) it is unlikely that they will be applied in the future unless psychologists suggest them. To what extent these conclusions apply to experts working in other domains is uncertain.

Our theoretical work has focused on the distinction between pattern-matching processes and processes involving functional relations. Hammond worked out a general theoretical framework that encompasses these two major concepts and shows how they are related to one another. This work is reported by K. R. Hammond, Judgment and decision making in dynamic tasks, soon to be published in Large Scale Systems.

## Reports

Hammond, K. R. (in press). Information models for intuitive and analytical cognition. In A. Sage (Ed.), Concise encyclopedia of information processing in systems and organizations. Oxford: Pergamon Press.

Hammond, K. R. (in press). Judgment and decision making in dynamic tasks. Large Scale Systems.

Lusk, C. M., Stewart, T. R., & Hammond, K. R. (1988). Judgment and decision making in dynamic tasks: The case of forecasting the microburst. (Technical Report No. 284). Boulder: University of Colorado, Center for Research on Judgment and Policy.

Lusk, C. M., Stewart, T. R., & Hammond, K. R. (1988). Toward the study of judgment and decision making in dynamic tasks: The case of forecasting the microburst. (Technical Report No. 273). Boulder: University of Colorado, Center for Research on Judgment and Policy.

Stewart, T. R., Moninger, W. R., Grassia, J., Brady, R. H., & Merrem, F. H. (1988). Analysis of expert judgment and skill in a hail forecasting experiment. (Technical Report No. 272). Boulder: University of Colorado, Center for Research on Judgment and Policy.

## Forthcoming Presentation

Potts, R., Lusk, C., Hammond, K., & Stewart, T. (1989). Expert judgment in the nowcasting of microbursts. Paper to be presented at the Third Annual International Conference on the Aviation Weather Systems, Anaheim, CA.

Appendix A

Study VI: A Laboratory Experiment: Precursor Assessment by Forecasters

Cynthia M. Lusk

Kenneth R. Hammond

May 1988

Center for Research on Judgment and Policy

Institute of Cognitive Science

University of Colorado, Boulder

The purpose of this report is to present the preliminary results of an experiment conducted to study the cognitive aspects of the nowcasting of microbursts. In particular, the report focuses on analyses of those data most pertinent to NCAR's immediate needs. In a previous report (Lusk, Stewart, and Hammond, 1988), we have outlined a hierarchical model depicting the steps between the storm environment and a judgment about microbursts, which is presented in Figure 1. The links between each phase in Figure 1 represent points at which forecasters' judgment processes are involved. One of our previous studies (Study III) had indicated some degree of disagreement among the forecasters regarding extraction of the precursor values from drawings of radar data and clouds. The present study was conducted to clarify those findings; here we use a situation more representative of that in which the forecasters normally operate.

## Procedure

The subjects in this experiment were four of the five microburst forecasters who participated in our previous studies.

The experiment was conducted in two phases. We began the first phase according to the procedure outlined in our 2 March 1988 research proposal and detailed below. After completion of two cases (one microburst and one null case), the study was halted for a preliminary assessment of the procedure and results. The agreement among the forecasters was found to be so low that a meeting was held to discuss whether further documentation of the conclusion would be cost beneficial. It was decided that further data would be worth acquiring and the experimental procedure was modified to collect those data. The procedures are detailed below.

### Overview

During each experimental session, the forecaster was seated in front of a large computer terminal used to present color Doppler radar displays. The experimenter was seated in front of a computer terminal that was used to run the experimental session. At the first session of each phase of the experiment, the forecasters were presented with instructions regarding how the experiment would proceed. The forecasters were presented with a volume of radar data, after which they made judgments of precursor values and the probability of a microburst. The presentation of data and making of judgments was repeated until completion of each case.

### The Cases

Six cases were used to generate the data in this report: two in the first phase and four in the second phase. Half of the cases in each phase were null cases and half were microburst cases.

Each case was arranged on a tape in consecutive volumes. Each volume consisted of 13 scans, starting with either the .5 or 1.1 scan and terminating with either the 34.8 or the 39.9 scan. In the first phase, Case 1 included six volumes. The data for Case 2 spanned eight volumes. However, one volume was skipped due to faulty data. In addition, one volume in Case 2 only included the lower seven scans. However, judgments were still collected for that short volume. In the second phase all cases included four volumes of data. Each case terminated before the microburst was evident on the lowest scan or before any obvious or substantial decrease in the intensity or height of the cell in the null cases.

## The Judgments

The forecasters were asked to make judgments of the six precursor values they had indicated to be the cues in Study I: descending core, collapsing storm, convergence/divergence above cloud base, convergence/divergence at or below cloud base, notch, and rotation. In addition, forecasters made judgments of the probability of a microburst occurring in the next 5 to 10 minutes.

The judgments regarding precursor values and probability of a microburst were made on the same scales as utilized in our previous studies. In addition, to the right of each rating scale was a blank for the forecasters to insert their confidence in their precursor judgments. The rating sheet is shown in Figure 2.

In the first phase, judgments were made after each volume.. Therefore, judgments were made six times for Case 1 and seven times for Case 2. In the second phase judgments were made after all but the first volume. Thus, three judgments were made for each of the four cases in the second phase.

## The Experimental Session

At the beginning of the first session in each phase, the forecasters were provided with instructions to read. For the first phase the instructions explained how the experimental sessions would proceed. They explained that each case consisted of several volume scans over time of a cell that did or did not produce a microburst, starting with the lowest scan at the earliest time. When they were finished observing each scan, the forecasters were instructed to tell the experimenter that they were ready for the next level scan. The forecasters were given up to thirty seconds to view each scan. After completion of a volume in this manner, the forecasters filled in the rating sheet. In addition, the instructions stated, in part:

> At the time of the first volume you can assume that a microburst is not presently occurring. Please assume before observing the first scan, that on the basis of prior information

(morning soundings, etc.) you have already reached the conclusion that the likelihood of a microburst on this day is .50. Then adjust your probabilities of a microburst after observing the radar data. Each case will terminate prior to evidence of outflow from a microburst or evidence that the storm is obviously dissipating.

Finally, the forecasters were given instructions to think aloud.

The instructions for the second phase explained the changes in the experimental procedure. The forecasters were informed that they would receive the noon sounding data, view only four volumes of data, and make judgments as in the first phase after the second through fourth volumes. In addition, the instructions explained that the scans would be presented continuously and that they would not need to think aloud.

The forecasters were provided with blank paper for taking notes and felt tip pens to mark the screen. The date for each case was masked on the computer screen. At the beginning of each case, the forecasters were told the coordinates where the cell they were to attend to was presently located.

Prior to presentation of each case in the second phase, the forecasters were presented with the eleven o'clock sounding for the day from which that case was drawn. The subjects were then asked what the probability of a microburst occurring was, based on the sounding information alone.

In the first phase, half of the forecasters were presented with Case 1 first, and half were presented with Case 2 first. In the second phase, the cases were arranged on a tape in a fixed order. Each forecaster began with a different case, but otherwise the order of presentation was fixed.

## Results and Discussion

### Verbal Protocols

Examples of the verbalizations are provided in the Appendix. Although no formal analyses have been completed on the verbal protocols collected during the experimental sessions, informal inspection indicates that during observation of the radar data the forecasters were primarily operating at Phases D and E in our hierarchical judgment model (see Figure 1). That is, the verbalizations primarily concern translating the radar data to information such as the maximum reflectivity values, convergence or divergence, and noting the occurrence of features such as a notch at each level scan. The dynamic nature of the task was evident in the verbalizations when the forecasters made comparisons of reflectivity or velocity features between different levels or times. Such comparisons may also indicate forecasters' integration of maximum reflectivity values over time and height into judgments of cue values such as descending core.

Preliminary review of the verbalizations made at the time of judgments yields little insight into the judgment process. For the most part, the forecasters provide a dichotomous yes or no value regarding the occurrence of each precursor, then decide exactly what value on the scale to circle. The cognitive process for making the probability of a microburst judgment was not evident. Apparently this takes place on an intuitive level. No calculations or applications of a principle or formula for organizing the information are evident in the protocols. This result makes the hypothesis that forecasters combine information in a linear additive fashion plausible.

Rating Sheets

The only analyses conducted to date concern the agreement between forecasters' judgments of precursor values and agreement between forecasters' judgments of the probability of a microburst. The data used in these analyses were the judgments made after each volume. Thus, 25 data points are possible for each subject (some volumes have a slightly lower number of data points in instances where forecasters did not provide ratings). The correlations between the judgments of each pair of forecasters were computed for each precursor and are presented in Tables 1 through 6. Similarly, the correlations between judgments of the probability of a microburst were computed and are presented in Table 7.

Tables 1 through 7 clearly indicate a lack of agreement between forecasters regarding both the precursor and probability judgments. Although many of the correlations were substantially larger than zero (and are, in fact, statistically significant), they are all substantially smaller than one or any other level of acceptable agreement.

Comparison of Tables 1 through 6 indicates a higher degree of agreement on some precursors than on others. Particularly noteworthy are the low and even negative (!) correlations for judgments of descending core. This result is particularly important because this precursor is the one which forecasters weighted most heavily in arriving at microburst probability judgments (as indicated in Study I). Our previous Study III also indicated some disagreement among forecasters, but not to the degree indicated in the present study. The present study is a much better indicator of the degree of disagreement given its representative design. Thus the representative conditions produced lower rather than higher agreement, in opposition to expectations. The higher agreement in Study III may be due in part to the fact that those judgments were made with clearly delineated schematic cloud and radar drawings, rather than actual radar data.

Agreement regarding precursor values was highest for the two convergence precursors, second highest for collapsing storm and notch, and lowest for rotation and descending core. Of course, future research will

need to address how agreement can be improved. It is possible that the different levels of agreement between precursors may be due to the different levels of abstraction or stages necessary to make judgments of the precursor values. For example, convergence is perhaps the precursor value most directly obtained (from the radar velocities). In contrast, the descending core judgment requires that the forecaster combine information about maximum reflectivity values over times and heights.

Note that there is one very high correlation regarding the probability of a microburst in Table 7: that between forecasters 1 and 4 of .88, a result that raises some interesting issues. First, note that although these two forecasters are in high agreement regarding the probability of a microburst, their agreement regarding the value for descending core is essentially zero (Table 1). In Study I, both of these forecasters gave the highest weight to descending core among the six precursors. In the present study they show no agreement regarding that cue value, yet they show high agreement regarding the probability of a microburst. Such a finding is puzzling and deserves a great deal of consideration. Possibly the forecasters are utilizing some other information in arriving at their microburst probability judgment. Second, other statistical analyses may yield insight into the discrepancy. For example, analyses run separately on the null and microburst cases show that the high correlation for probability of a microburst judgment is to a large extent due to agreement on the microburst cases ($r = .95$), rather than agreement on the null cases ($r = .47$). However, a striking result is that for both the microburst and null cases, the correlations between the forecasters' judgments of descending core are negative ($r = -.38$ for microburst cases, $r = -.18$ for null cases). Similar comparisons may be made for other forecasters and judgments.

## Conclusions and Implications

The most important and clear cut finding from these preliminary analyses is a pervasive lack of agreement among the forecasters' judgments of precursor values. Although in many cases the level of agreement is at a moderate level, it is important to note, as we have previously indicated, that the level of measurement at any level in the judgment process (see Figure 1) sets the upper level for accuracy at the final stage of microburst prediction.

We have demonstrated how the analysis of the cognitive aspects of forecasting can help delineate the judgment process and potential sources of error. Continued application of this approach would be helpful for improving agreement (and possibly accuracy). A first step may be to use our hierarchical approach in decomposing the precursor judgments into their components, much as we have in Figure 1. That is, one would want to delineate what features in the radar data are cues for each precursor and how those features are combined. Such an exercise would prove useful for

both clarifying the definition of each precursor and for training others in the use of Doppler radar for microburst forecasting.

Appendix

## Example Protocols for Study VI

Subject 1
Case I

S: Okay where are we? The number of the next volume, 4. Ah what's happening? Uh very weak divergent flow at the surface, very weak, only three meters per second. And we've got about 55. It's 55. Very weak. Huh again we're we see at these 55, we get divergence again above. See it really looks like we're getting a little, it's diverging out above cold air, but it's weak. And it gone, oh wow. We get some actualy 60 this time, reflectivity. A lot more reflectivity. And actually we're showing a little convergence now. Oh wow it's up to 60 now. But velocity feature not very strong, slight. Still 60, no good velocity feature. I'm not wild about the angle we're getting now. If there were convergence in that core we wouldn't see it well. Now at 55, I'll call it now, it's just only a touch of 60. Slight indication of that notch is at this level, now. This is 15 6 [pause] there's xxxx convergence into that too, hmm. Nice notch now, reflectivity 55. Can't see an obvious velocity feature with it though. Here's where we get the convergence. 45, 45 convergence. Okay we've lost a lot of reflectivity now. And we, now we're actually divergence. It's slipping down into the about 45, maybe 40, at 30 degrees. Oh it's gone only 25 left so we have a real collapsing case here. Boy that was faster wasn't it.

E: yeah

S: I had to move. Just trying to see xxx [silence] The top's coming down. Okay now uh descending reflectivity core, yeah it's still, it's not one of the obvious, the most obvious cases in the world, but it's still descending. I'll put a 7, confidence is only about 50 percent. Collapsing storm, it is collapsing but it's not the most obvious one you ever saw. So I'll put 7, confidence at 60 percent. Organized convergence above cloud base, yes it's still there. It's still, and it's actually descended slightly with time I see. Not much, it's still, it's still primarily in the three to four kilometer zone which is a good zone for it. It's not that strong and organized. I put confidence only at 60, meaning I don't think it's all that significant. Organized, there's still a divergence below cloud base, and I really think that may be significant. Um I'm going to put, I'm circling the one and two, saying, I'm putting 70 on it cause I think the outflow is really divergent above the cold air. It may not make it to the surface very strong. Good reflectivity notch now between 2 and 3 kilometers. I'll put a 9 on it, confidence is, well it's there, 90. Rotation was um not as good, it was weak. Last time I think I had weak. I xxx put down a 6. Um confidence is only 50 percent. Okay now if we're going to have a microburst that's going to occur in this period, I'm not

very, I think it's only going to be a very weak outflow though cause the reasons I've given. Last time I gave 25 percent. I'll go with 30 and hope I'm right.

Subject 2
Case 2

S: [Silence] Okay max reflectivity here is 55. Still got weak convergence delta V is 3, okay. [Pause] 55 again, two point two. Weak convergence again. Okay. xxx don't see it this time. 4 and a half degrees, 55. Um still convergent weakly delta V is 3, okay. [Pause] 6.7 xxx 55. [Silence] Um not much going on that's really different, okay. 8.8 is 55. 55 [pause] hmm. A suggestion of xxx divergence on the north edge of cell, delta V is about 3. It's still pretty weak, okay. [Silence] 50 DBZ, 11 degrees. Got that wind change xxx, okay. [Pause] 50 DBZ again. [Silence] Okay. [Silence] Well that's interesting, huh. 50 DBZ, ??erosion?? echo in the back. Notch is still there. It's kind of filling in though, there's mid-line with more echo to the west of the cell than there has been previously. [Silence] Cyclonic, anti-cyclonic couplet there. Um okay [silence] 50 DBZ, this storm really is tilted in height. Sort of see convergence xxxx weak xxxxx [silence] okay. [Silence] okay xxx DBZ [silence] There's some shear areas but nothing really significant. This is 22 degrees, um [pause] okay. [Silence] 45, again we've gotten a couple of shear areas. Cyclonic, anti-cyclonic shear not real couples to speak of [pause] okay. [Silence] xxx xxxx [pause] cyclonic, anti-cyclonic shear okay. [Pause] The cell's falling apart xxx. 35 DBZ. There's convergence ??in the anvil??. [mumbles] 6. [silence]

S: Uh reflectivities are still maintaining themselves pretty well. [Silence] Slightly increasing aloft and then decreasing at the very highest angle. So we don't have a descending core. And it's not collapsing. There's no real convergence above cloud base, except in the xxxx. [Silence] Um [silence] there's not, there's convergence at or below cloud base. xxxx xxxx kilometer, there's that one little spot of divergence ??at one kilometer?? It's really weak though. [Silence] The notch has become weaker. Not as well defined. And there's also xxx flow xxx so I'm going to rotation, no there's some cyclonic shear and that's it. Probability of a microburst within the next 5 to 10 minutes, I'm still going to stick with the 50 percent.

Subject 4
Case 1

S: xxx you look at that point 5 degree velocity and there's nothing there. There is not a microburst outflow. There's some garbage right there, but that's not real. And uh looking all the way up at 2.2 we don't

really see any divergence or velocity structure. And we've got the high reflectivity xxx so unless we see some dramatic increases in velocity structure, which we don't really see here at 4.5, it's going to be awful hard to say yes we're going to get something. And uh even at 6.7 we're not seeing any good strong velocity features associated with that core. [Long Silence] slight hint that there may be convergence coming in here that we can't see associated with that notch. And xxx interesting to look at it from a from a radar out here where we could get a better view. Still seeing that notch, but again, as I say, it's not that good of a velocity structure. I did see some sign of convergence xxx. [Silence] Saw some. xxx [silence] xxx rotation xxx some convergence not really that good. [Silence] xxx looks about the same as it was before [silence] okay.

S: xxx other sheet xxx put down thing xxx for can't remember for sure. ??We do have?? some descent of core. The storm has collapsed already. I think there's a slight xxx still kind of collapsing. Uh xxx not really much happening above cloud base xxx. xxx Part of why you think collapsing storm. Slight indication of xxx. We got an indication xxx notch. Well nothing happened last time. Still not seeing it, we've got the high reflectivity down so, not willing to say no chance anymore, but uh got to start backing off a little bit on that probability. I'll be a little less convinced that something's going to happpen.

<u>Subject 5</u>
<u>Case 2</u>

S: Okay. It takes it forever. Oh we're going to start with point 5. [Silence] Oh yeah, this guy's racing off to the north, and 55 DBZ core. [pause] And a little convergent shear line still with us way off to the south. Oh that's what happened to the cell. It moved off of its convergent line. Now it's lost its low level support. It's going to crash, okay. [Silence] Oh that's why the core crashed down in such a hurry. [Silence] That's right I did see a sizeable increase in reflectivity. And that's what happened to it. Okay.

E: Is that an okay for me?

S: Yeah, that's an okay okay. [Silence] Oh gosh 60 DBZ. [Silence] No velocity features at all associated with the cell at 4.5 degrees. [Silence] Surprised it hasn't put out an outflow, okay. Wonder why not? [Silence] Oh gee, everything's back down to 55 DBZ now. [Silence] huh. Still no real velocity features. It's really just a flat field. Okay. Notch on the side. [Silence] huh let's see, not much at all going on. Strange, we're up at 8.8 degrees and I don't see much of anything, huh. Okay, go to the next one, if you haven't already. [Silence] 50 to 55, well a little bit of cyclonic shear. Certainly a notch. Okay. [Silence] Oh another cyclonic shear right in the middle of the cell. [Silence] Oh yeah,

a little bit of convergence right there, okay. [Silence] Oh hurry up [silence] yep, a little bit of convergence now in the middle of the cell. [Silence] Okay. [Silence] Oh rotation hanging off, way off on the end out in the area of no, not much signal. Uh now we're seeing convergence peppered about here, hither in the thither. Rotation down in the south end where we've always seen it. xxx okay. [Silence] Oh there's a clear rotation near that notch, cyclonic rotation. Okay. [Silence] huh a little bit of divergence right up here. 25.8 degrees, cyclonic shear to the south, probably strong rotation. Huh. [Silence] Is it doing anything? [referring to computer] [Silence] Oh yeah now I see cnvergence on the western end, right where that notch, okay. [Silence] Oh there's convergence all over the place, 34.8. Uh max reflectivities xxx 40 to 45. Okay.

E: That's it on that one.

S: Okay. Descending reflectivity core, it's obvious. Collapsing storm, probably is, but not real sure yet. Organized convergence or divergence above cloud base, you betcha. Not much convergence at or below cloud base, I didn't seen anything. And I'm pretty sure I didn't see anything. There's a reflectivity notch. There's rotation. I'm a little concerned that I didn't see any divergence at the surface, but what the heck. 90 percent, or is this [silence]

## Table 1

### Correlations Among Judgments of Descending Core

| | F1 | F2 | F4 |
|---|---|---|---|
| F2 | .14 | | |
| F4 | -.06 | .12 | |
| F5 | .10 | .35 | -.14 |

## Table 2

### Correlations Among Judgments of Collapsing Storm

| | F1 | F2 | F4 |
|---|---|---|---|
| F2 | .69 | | |
| F4 | .47 | .53 | |
| F5 | .57 | .40 | .17 |

## Table 3

### Correlations Among Judgments of Convergence Above Cloud Base

| | F1 | F2 | F4 |
|---|---|---|---|
| F2 | .65 | | |
| F4 | .71 | .49 | |
| F5 | .58 | .53 | .45 |

## Table 4

### Correlations Among Judgments of Convergence at/or Below Cloud Base

| | F1 | F2 | F4 |
|---|---|---|---|
| F2 | .54 | | |
| F4 | .43 | .76 | |
| F5 | .77 | .59 | .45 |

## Table 5

**Correlations Among Judgments of Notch**

| | F1 | F2 | F4 |
|---|---|---|---|
| F2 | .38 | | |
| F4 | .51 | .25 | |
| F5 | .61 | .57 | .34 |

## Table 6

**Correlations Among Judgments of Rotation**

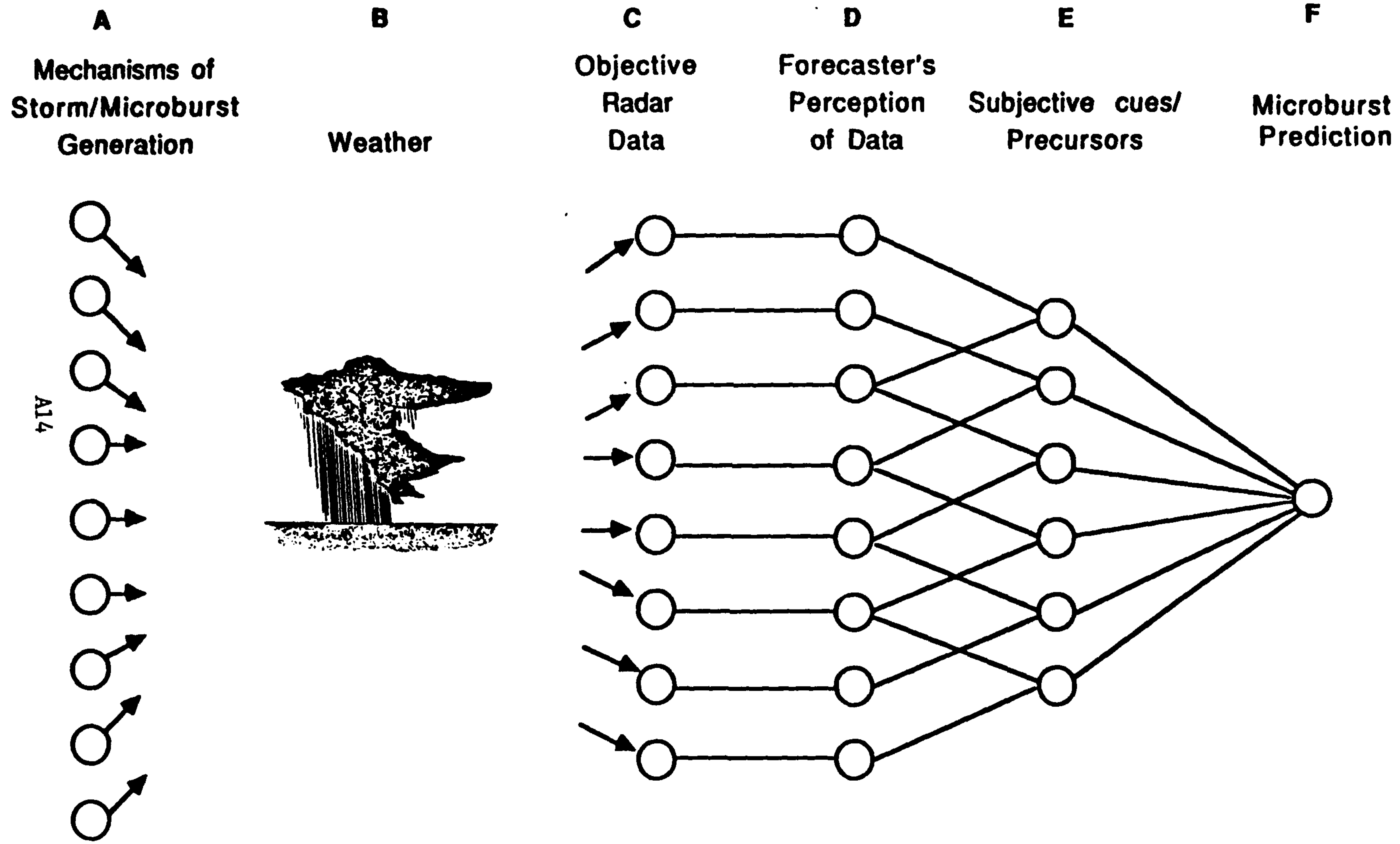
| | F1 | F2 | F4 |
|---|---|---|---|
| F2 | .06 | | |
| F4 | .12 | .39 | |
| F5 | .51 | -.01 | .26 |

## Table 7

**Correlations Among Judgments of Probability of a Microburst**

| | F1 | F2 | F4 |
|---|---|---|---|
| F2 | .60 | | |
| F4 | .88 | .45 | |
| F5 | .31 | .15 | .19 |

Figure 1
Sequence of Phases in Microburst Forecasting



## Figure 2: Precursor Judgment Scales

### RATING SHEET

| | | | CONFIDENCE (%) |
|---|---|---|---|
| DESCENDING REFLECTIVITY CORE | 1 2 3 4 5 6 7 8 9 10 | NOT DESCENDING — QUESTIONABLE — OBVIOUS | ________ |
| COLLAPSING STORM | 1 2 3 4 5 6 7 8 9 10 | NOT COLLAPSING — QUESTIONABLE — OBVIOUS | ________ |
| ORGANIZED CONVERGENCE (DIVERGENCE) ABOVE CLOUD BASE | 2 1 0 -1 -2 -3 -4 -5 -6 -7 -8 | DIVERGENCE — LOW CONVERGENCE ($10 ms^{-1}$ /5km) — HIGH CONVERGENCE ($40 ms^{-1}$ /5km) | ________ |
| ORGANIZED CONVERGENCE (DIVERGENCE) AT OR BELOW CLOUD BASE | 2 1 0 -1 -2 -3 -4 -5 -6 -7 -8 | DIVERGENCE — LOW CONVERGENCE ($10 ms^{-1}$ /5km) — HIGH CONVERGENCE ($40 ms^{-1}$ /5km) | ________ |
| REFLECTIVITY NOTCH | 1 2 3 4 5 6 7 8 9 10 | NO NOTCH — QUESTIONABLE — OBVIOUS | ________ |
| ROTATION | 1 2 3 4 5 6 7 8 9 10 | NONE — AMBIGUOUS — WEAK — TIGHT | ________ |

**probability (0.0 - 1.0) of microburst within 5 - 10 minutes:** ________

**PLEASE LIST ON THE BACK OF THIS PAGE ANY OTHER FACTORS YOU CONSIDERED WHEN MAKING YOUR MICROBURST JUDGMENT**

Appendix B

# Cognitive Aspects of Forecasting the Microburst: Research Results, Conclusions, and Recommendations

Kenneth R. Hammond

May 1988

Center for Research on Judgment and Policy
Institute of Cognitive Science
University of Colorado, Boulder

## Table of Contents

# Part I

## Executive Summary

### Resumé of Results

The role of the judgment and decision psychologists in the microburst project has been to study the cognitive processes of research meteorologists attempting to forecast (nowcast) microburst events. The goal of the research is to assist in the improvement of such forecasts. Six studies have been carried out since April 1987. Study I found that: (a) there was only modest agreement between and within forecasters in forecasts based on error-free precursor profiles (a best-case scenario), (b) a linear model of the judgment process was a good predictor of forecaster judgments, and (c) it was a better predictor than the process reported by the forecasters. Study II also found only modest agreement among forecasters when they were asked to group together similar precursor profiles and to construct storm images from them. Study III examined the reverse process--constructing precursor profiles from the forecasters' storm images--and found only moderately accurate reproductions. Study IV found that although the 1987 Buckley field experience led to improved agreement among forecaster judgments, agreement regarding similarity judgments of precursor profiles remained modest. To summarize, results from Studies I-IV indicate that the human forecasting process is far from being a unified one and far from being a consistent process.

Because of the persistent finding of only modest agreement (i.e., modest inter- and intra-observer consistency) among forecasters we next investigated the accuracy of an algorithm that capitalized on providing entirely consistent (i.e., same data, same response) judgments under conditions that formally simulate the forecasting situation (even though a multivariate analysis of precursor-microburst (truth) data would not be available). By comparing the effect of perfect consistency with perfect model accuracy it would be possible to discover what could be gained by improved consistency relative to improved scientific understanding. Therefore, Study V was constructed to compare the accuracy of (a) a scientifically ignorant forecasting equation consistently applied with (b) a perfectly accurate conceptual model of microburst events consistently applied under sixteen different conditions, each of which contained 100 different storms. It was found that under the most realistic assumptions (moderate intercorrelations among precursors, moderate environmental uncertainty) there would be little difference in accuracy between a scientifically ignorant forecasting equation and a perfectly correct model, both consistently applied. Thus, the results indicate that little is likely to be gained by improving the conceptual model; more is apt to be gained by reducing uncertainty in other steps in the forecasting process. The question then became: Where does the uncertainty lie, and what can be done to reduce it?

In order to answer these questions Study VI investigated forecaster judgments based on actual Doppler radar displays of six cases. Study VI found low to moderate agreement among forecasters in the judgments of precursor values--particularly in judgments of "descending core"--the precursor given the most weight both by the forecasters empirically in Studies I-IV, and in discussions. For example, agreement between forecasters in their assessments of "descending core" ranged from -.14 to .35 (in correlation coefficients). Two forecasters agreed very closely (.87) over the six cases in their judgments of the likelihood of a microburst, yet disagreed completely (-.06) in their judgments regarding descending core. Whatever is causing these two forecasters to agree in their judgments of microburst probability, it is not because they agree on the presence or absence of a descending core, a result that is inexplicable from verbal explanations of the forecasting process. Therefore the answer to the above question--Where does the uncertainty lie?--is that it lies in precursor judgments, if nowhere else. (Caution: These results are based only on six cases.) The answer to the question--What can be done to reduce uncertainty?--is presented under "Recommendations" below.

Note. It is our understanding, based on conversations with Steve Campbell (4 April 1988), that the MIT computer program for detecting the presence and location of "descending cores" is heavily dependent upon the declaration by one meteorologist that a descending core is or is not present. Therefore, Steve Campbell should be advised of the results of Study VI.

## Conclusions

1. Taken together, the results of Studies I, II, III, IV, and VI indicate that forecasters' predictive accuracy for microburst events and null events must be low. In particular, the results of Study VI indicated that the perceptual assessments of precursor conditions by the forecasters are a major source of error. The results of Study V, together with those of Study VI, indicate that although improvement in the conceptual model is not likely to aid matters much, improvement in the accuracy of precursor assessment will.

2. Whatever the value of the Roberts-Wilson conceptual model might be for understanding microbursts, the model is of little practical value for predicting the occurrence or nonoccurrence of microbursts. (This circumstance is not unusual; understanding and prediction are not always closely linked.) It is clearly possible that the difficulty is almost certainly due to the difficulties of information processing, rather than a lack of meteorological knowledge. For as matters stand now, forecasters must incorporate what, by any standard, is a great deal of changing, ambiguous information of uncertain value without the necessary cognitive supports. They must make many (50-100?) perceptual judgments over many different volume scans regarding both reflectivity and velocity readings at

the same time that they are organizing data from these measurements into an overall judgment of microburst likelihood, all of this without explicit definitions of such important precursors as descending core, and explicit methods or principles for combining the data created by their perceptual judgments. Accurate forecasts under these conditions would, in fact, be surprising.

3. The perceptual conditions of the Doppler display do not favor accurate assessments of precursors. The perceptual environment provided by this flat, rectangular, two-dimensional array of numerous color contours is, on the one hand, impoverished; it does not provide the human visual perceptual system with the rich three-dimensional display of objects in textured space for which the visual system is so well adapted and for which it is so effective. On the other hand, the Doppler radar is not lean enough to provide the unambiguous "pointer-readings" of, for example, the cockpit instrument. Furthermore, present conditions require that perceptual judgments of both velocities and reflectivities must be made over several volume scans, a cognitive activity that makes severe demands on memory. In short, the information display conditions are not conducive to accurate perceptual judgments nor to the integration of perceptual data into a scientifically based judgment.

4. If these conclusions are correct, and we see no hard evidence that would challenge them, it is only reasonable to raise the question of how best to remedy the situation. Although the MIT research team is currently attempting to develop a computer scanning technique that will identify precursors, "descending core" in particular, the procedure, as noted above, is dependent upon the perceptual judgments of descending cores made by a single meteorologist. Given the results of Study VI, these results are not likely to be replicable. Therefore, measurement of precursors remains tied to the perceptual judgments of forecasters. That brings us to the second of the two questions above: How can human precursor judgments be improved?

Recommendation 1: Create Rigorous, Explicit Definitions of Precursors

We recommend this procedure for reasons indicated below.

First, recall that: (a) Studies I-V show that increased refinement of the Roberts-Wilson conceptual model offers little promise for improved forecasting/nowcasting, and (b) the specific problem to be faced is lack of agreement among expert research meteorologists in their perceptual judgments of precursor events. Improvement in precursor measurement can lead to more accurate forecasts, even if a scientifically ignorant forecasting equation is used. Without such improvement, even a perfectly accurate conceptual model will be of little value.

Second, note that at present there appear to be no written-out, explicit, agreed-upon definitions of conditions that indicate the presence of precursors. This even appears to be true of "descending core." (We use the word "appear" because we cannot be certain that such definitions do not exist; but we have not seen them, nor have any of the forecasters referred us to them, and at least one meteorologist reports that there are none.)

Therefore, if

1. the descriptions of the perceptual data measurements and the empirical results of the above studies (particularly Study VI) are true,

2. it is true that there is no set of explicit definitions or instructions for the identification of precursors,

then we recommend that steps be taken to remedy the omissions described in 1. and 2. above, and we offer the following suggestions regarding the procedure for construction of definitions.

These should follow from the best scientific theoretical base available. These theoretical definitions should be translated into observables by means of a public (i.e., use of more than one expert) critique. The procedure might well involve schematic, pictorial, and actual Doppler radar pictorial images. Empirical tests of actual agreement on <u>definitions</u> (not judgments) should be employed, rather than relying on consensus based verbal expression of agreement.

<u>Recommendation 2: Create a Formal Training Program for the Identification and Assessment of Precursors</u>

There appears to be no formal training and evaluation procedure for the judgment of precursors. (Again, we use the word "appears" for the same reasons as above.) Of course, we realize that the forecasters studied here have spent perhaps thousands of hours observing Doppler displays. Nevertheless, such experience by itself cannot substitute for formal training exercises that track performance and provide feedback.

<u>Recommendation 3: Carry Out a Field Performance (or Close Simulation Thereof) Test</u>

It cannot be taken for granted that once agreement on theoretical and operational definitions of precursors has been established, and training has brought perceptual judgments up to desired agreement levels, that agreement on perceptual judgments of the changing events of actual cases will follow. Empirical tests should be used (as in Study VI) to determine the degree of agreement under close simulation of working conditions. Such tests are essential because it is presently impossible to establish

accuracy by comparing precursor judgments of, say, descending core with independent objective, readily verifiable data. Training might well be carried out in cooperation with psychologists with experience in improving judgments under uncertainty.

To sum up: Three recommendations are provided for ways to improve predictions of microburst events on the basis of present knowledge afforded by the Roberts-Wilson model and other scientifically-based information: These are (a) clarification of definitions, (b) training, and (c) simulated field testing.

## Further Implications of Studies I-VI

### 1. Implications Concerning Multivariate Analysis of Relations Between Precursors and Microbursts and Null Events

The results of Study VI hold considerable significance for the evaluation of empirical relations between precursors and microburst and null events. Because critically important precursors such as "descending core" can only be identified by forecasters' perceptual judgments, and because these judgments vary considerably among forecasters, there is no possibility of measuring the empirical relation between judgments of precursor events and microbursts and null events, as matters stand now. If these theoretical relations cannot be tested empirically then the value of the enormous amount of data and precursor judgments already recorded must be called into question. Apparently the only empirical multivariate analysis of physical events that can be done would involve completely objective data (velocities, reflectivities). This would be a huge undertaking with results of doubtful utility because of the size of the data set. But this study should be given serious consideration.

### 2. Implications for Training Other Meteorologists

If there is little agreement among the NCAR meteorologists in their judgments of precursors, then training of other meteorologists will depend heavily on which meterologist is the trainer. We assume that this circumstance is undesirable.

### 3. Implications for Future Research

Apparently Study VI is one of the first attempts to carry out a study of forecasting using experimental, laboratory-type methods. Therefore we point out that the existence of considerable amounts of Doppler radar tape means that it is now possible to conduct true experiments that will permit the examination of (a) the relative efficacy of various forecasting methods, (b) the relative utility of various aids for forecasters, and (c) the relative advantages of various display methods and equipment (e.g., the three-dimensional display) as well as (d) the cognitive aspects of

forecasting. In short, the new ability to carry out experiments may offer meteorologists new opportunities, particularly if experiments similar to those in Studies VI and VII can be carried out prior to field studies. Such experiments would be highly cost effective relative to field tests such as those conducted during 1987.

## Part II

## Abstracts of Studies

### Study I: Agreement Among Forecasters Under a "Best-Case" Scenario

In Study I the meteorologists judged the likelihood of microbursts after observing precursor profiles of storm cells. The precursor profiles provided perfectly reliable observations because the forecasters were provided with exact precursor cue values rather than having to measure the precursor values perceptually. Thus Study 1 provided a "best-case" scenario. (It is a best case scenario because if the forecasters were required to judge the values of the precursors, at least some error would be introduced in the forecasting process thus lowering the agreement among the forecasters. Additionally, every forecaster thus saw exactly the same precursor values.) Results indicated that the forecasters' judgment process is adequately represented by a linear model. A nonlinear model of the forecasters' cognitive process, which the forecasters claim to use, failed to reproduce the forecasts as well as the linear model. It is important to note that only modest agreement was found among the forecasters' judgments in this "best-case" scenario.

## Study II: Can Storm Images Be Constructed from Precursor Profiles?

Study II tested the meaningfulness of Study I by investigating whether the precursor profiles readily evoked pictorial images of storms. This was done by asking the meteorologists to sort the precursor profiles into similar categories and then to draw pictures of the storms the categories represented. Results indicated that images were readily evoked by the precursor profiles, thus confirming the meaningfulness of the profiles used in Study I. In addition, the forecasters' sortings provided independent confirmation of the linear model of information integration found in Study I (i.e., the linear model for each forecaster predicted his/her sorting of the profiles). Agreement among forecasters with regard to the sorting of profiles was found to be modest, however, thus suggesting that the same error-free precursor values may give rise to different storm images for different forecasters.

## Study III: Can Precursor Profiles Be Constructed from Storm Images?

In Study II we asked whether storm images could be constructed from precursor profiles grouped by each forecaster. In Study III we investigated the reverse process; can forecasters construct precursor values from an examination of storm images? To what extent will forecasters agree on the precursor values when they observe the same storm images (of both natural and radar form)? Forecasters readily constructed precursor profiles, but some precursor values (descending core) were more accurately predicted than others. Agreement among forecasters with regard to precursor values based on their observation of the storm drawings was only modest.

## Study IV: Effects of the 1987 Field Experience on Stability of the Judgment Process

The effect of field experience on the persistence of the linear model as a representation of the judgment process and agreement among meteorologists was investigated by asking meteorologists once again to sort the precursor profiles used in Study I into meaningful categories. The linear model again was found to predict the sorting, thus confirming the results obtained in Studies I and II. Agreement among the meteorologists was found to be somewhat higher after the field experience, although considerable disagreement remained.

## Study V: Evaluation of the Forecasting Accuracy of Scientifically Ignorant Forecasting Equations Relative to the Accuracy of Perfectly Accurate Prediction Models

Because the ultimate aim of the microburst research project is to develop an algorithm for nowcasting microbursts, we investigated the question of whether it is necessary to develop a sophisticated, scientifically informed algorithm to do that. Could a simple, scientifically and empirically ignorant forecasting could be used instead? To answer this question we created sixteen sets of 100 different storms each. The 16 sets were created from (a) two different storm models (one complex, one simple), (b) using two different levels of intercorrelation (one zero, one moderately high), and (c) four levels of uncertainty. Two different forecasting equations (one complex, one simple) were applied to the 1600 storms thus developed and the relative accuracy of each was evaluated quantitatively.

The results were consistent with those from previous research. The simple, incorrect additive equal weight forecasting equation was as accurate in terms of both hit rates and correlation coefficients as the complex, correct algorithm when (a) precursors were intercorrelated and (b) uncertainty was at least moderate. This result argues that a simple forecasting equation should be tested with actual microburst data that includes ground truth as soon as possible. Should that test confirm the simulation test (as we believe it will), plans should be made to test such an equation in the field in 1988. Perhaps a simple, low cost algorithm that will meet FAA standards of accuracy is already available. If so, considerable time, energy, and money might be saved.

## Study VI: A Laboratory Experiment: Precursor Assessment by Forecasters

The results of Studies I-V cast doubt on the efficacy of judgments of the occurrence of microbursts. All five studies, however, suffered from a lack of representativeness; that is, the forecasters were *not* presented with the actual circumstances in which microburst forecasts are made. Study VI was designed to remedy that situation. Rod Potts retrieved microburst and null cases from the 1987 field study. Thus, accuracy of forecaster judgments under representative conditions could be ascertained for the first time, as well as agreement on judgments of precursor values. Present conclusions are based on results obtained from three null cases and three microburst cases.

If inter-observer agreement is taken as a measure of intra-observer agreement, then it is clear that forecasting accuracy cannot reach even modest levels, even if perfect conceptual models were to be developed. Given the results of Studies I through VI there seems to be little justification for trying to develop better conceptual models. Results to date argue for focusing first on ascertaining the ecological validity of the descending core precursor; that is, what is the empirical relation between the best available measurement of descending core and the appearance or nonappearance of microbursts? To the best of our knowledge, very little data are available for ascertaining the answer to this question (although Rod Potts is collecting such data). A more general question is that of how best to assist forecasters in their efforts to identify precursor conditions displayed on the Doppler radar screen.

Appendix C

# Further Analyses of Expert's Judgments in a Dynamic Task

Kenneth R. Hammond

September 1988

Center for Research on Judgment and Policy
Institute of Cognitive Science
University of Colorado, Boulder

Dynamic Tasks
Kenneth R. Hammond

To investigate (a) the impact of information on judgments at specific times and (b) the change in judgments over time, we graphed each forecaster's judgments regarding the probability of a microburst as a function of each (volume) scan. These graphs are presented in Figure A1. Inspection of Figure A1 yields several interesting observations. First, there are wide individual differences in forecaster's predictions. In addition, the subjects did not converge on a similar judgment with the accumulation of more evidence; with a minor exception, accumulation of evidence has little effect on agreement. Second, judgments change very little over time; the lines joining subsequent judgments are nearly flat. Means were computed for the difference between each consecutive probability judgment for each forecaster and they are as follows: .06, .08, .10, .13. Means were also computed for the difference between the first and last judgments in a given case for each forecaster and they are as follows: .09, .13, .06, .24. In short, the forecasters change their probability judgments by only about ten percent on average. A oneway analysis of variance performed for each forecaster separately yielded no statistically significant differences in probability judgments due to new information for any of the forecasters. This is a surprising and important finding for two reasons: (a) the forecasters believe that their judgments are, in fact, influenced by incoming information; (b) forecasts may be as accurate when made early in the forecast process as when made with much more information at a later time. Taken together, these conclusions suggest that forecasting may be less problematic than NCAR currently believes it to be. Finally, it should be noted that the study would not have been carried out, and, therefore, this information would not have been obtained, without the participation of psychologists. One implication that ARI might draw from this is that given the new technological capacity to run dynamic situations repeatedly with different experts, considerable information may be obtained about judgment and decision making by experts in dynamic tasks by researchers in the field of judgment and decision making.

# FIGURE 1A

## JUDGED MICROBURST PROBABILITIES OVER TIME